Sequence Information from 42−108-mer DNAs (Complete for a 50-mer) by Tandem Mass Spectrometry

Daniel P. Little,† David J. Aaserud, Gary A. Valaskovic, and Fred W. McLafferty*

Contribution from the Department of Chemistry, Baker Laboratory, Cornell University, Ithaca, New York 14853-1301

Received October 2, 1995. Revised Manuscript Received May 8, 1996[⊗]

Abstract: Complete sequence information for an "unknown" 50-mer DNA and extensive sequence verification for another 50-mer and 42-, 51-, 55-, 60-, 72-, 100-, and 108-mer DNAs is obtained by electrospray ionization/Fourier transform mass spectrometry that supplies 10−100× higher accuracy and resolving-power data using nozzle-skimmer (NS), collisionally activated, and infrared multiphoton dissociation (IRMPD). In addition to the previously recognized 3′- and 5′-terminal (w and a) ions, internal ions (i) and MS/MS/MS of fragment ions provide unique structural information across the DNA. NS dissociation can also yield other new backbone cleavages (forming b, c, d, and r ions) that provide extensive 5′-end information. These spectra indicate that loss of the base T rarely triggers formation of w, a, or i fragment ions, a correlation of further sequencing utility. Point mutation screening is demonstrated using a modified 50-mer unknown; a 9.04 (theory 9.01) decrease in the molecular weight (M_r) value indicates A → T, while three IRMPD fragment ions pinpoint this mutation at base 27. Introduction (measurement time <1 min) of 8×10^{-16} mol of the 50-mer gave an M_r value with only a 0.2 error.

Introduction

Gene-level diagnosis is now highly promising. Genetic defect loci have been determined for such ailments as breast[1] and colon[2] cancer, Huntington's disease,[3] ataxia telangiectasia (AT),[4] cystic fibrosis,[5] and myotonic dystrophy;[6] more than 50 tumor types are due to mutations in the "cancer gene" that codes for the tumor suppressor protein p53.[7] Diagnoses, as well as identification of new defects, employ established methods of molecular biology[8] (e.g., separation by polyacrylamide gel electrophoresis, PAGE).[9] However, these have specificity problems such as insensitivity to base modifications; base methylation has been implicated in carcinomas.[10] Accurate, fast verification could also be critical in the time requirements for genome sequencing.[11] For verification as well as diagnosis, a proposed complementary method is characterization of an isolated definitive region(s) of the nucleotide using mass spectrometry (MS),[12−27] as MS is far more sensitive[28] and faster than PAGE. Although finding the expected molecular weight

† Current address: SEQUENOM Instruments GmbH, Mendelssohnstr. 15D, 22761 Hamburg, Germany.

⊗ Abstract published in *Advance ACS Abstracts,* September 15, 1996.

(1) Hall, J. M.; Lee, M. K.; Newman, B.; Morrow, J. E.; Anderson, L. A.; Huey, B.; King, M.-C. *Science* **1990**, *250*, 1684−1689. Wooster, R.; Neuhausen, S. L.; Mangion, J.; Quirk, Y.; Ford, D.; Collins, N.; Nguyen, K.; Seal, S.; Tran, T.; Averill, D.; Fields, P.; Marshall, G.; Narod, S.; Lenior, G. M.; Lynch, H.; Feunteun, J.; Devilee, P.; Cornelisse, C. J.; Menko, F. H.; Daly, P. A.; Ormiston, W.; McManus, R.; Pye, C.; Lewis, C. M.; Cannon-Albright, L. A.; Peto, J.; Ponder, B. A. J.; Skolnick, M. H.; Easton, D. F.; Goldgar, D. E.; Stratton, M. R. *Science* **1994**, *265*, 2088−2090.

(2) Nicolaides, N. C.; Papadopoulos, N.; Liu, B.; Wei, Y.-F.; Carter, K. C.; Ruben, S. M.; Rosen, C. A.; Haseltine, W. A.; Fleischmann, R. D.; Fraser, C. M.; Adams, M. D.; Venter, J. C.; Dunlop, M. G.; Hamilton, S. R.; Petersen, G. M.; de la Chapelle, A.; Vogelstein, B.; Kinzler, K. W. *Nature* **1994**, *371*, 75−80.

(3) Duyao, M. P.; Auerbach, A. B.; Ryan, A.; Persichetti, F.; Barnes, G. T.; McNeil, S. M.; Ge, P.; Vonsattel, J. P.; Gusella, J. F.; Joyner, A. L.; MacDonald, M. E. *Science* **1995**, *269*, 407−410.

(4) Savitsky, K.; Bar-Shira, A.; Gilad, S.; Rotman, G.; Ziv, Y.; Vanagaite, L.; Tagle, D. A.; Smith, S.; Uziel, T.; Sfez, S.; Ashkenazi, M.; Pecker, I.; Frydman, M.; Harnik, R.; Patanjali, S. R.; Simmons, A.; Clines, G. A.; Sartiel, A.; Gatti, R. A.; Chessa, L.; Sanal, O.; Lavin, M. F.; Jaspers, N. G. J.; Taylor, A. M. R.; Arlett, C. F.; Miki, T.; Weissman, S. M.; Lovett, M.; Collins, F. S.; Shiloh, Y. *Science* **1995**, *268*, 1749−1753.

(5) Hastbacka, J.; de la Chapelle, A.; Mahtani, M. M.; Clines, G.; Reeve-Daly, M. P.; Daly, M.; Hamilton, B. A.; Kusumi, K.; Trivedi, B.; Weaver, A.; Coloma, A.; Lovett, M.; Buckler, A.; Kaitila, I.; Lander, E. S. *Cell* **1994**, *78*, 1073−1087.

(6) Weber, J. L.; Wong, C. *Hum. Mol. Genet.* **1993**, *2*, 1123−1128.

(7) Harris, C. C. *Science* **1993**, *262*, 1980−1981. Culotta, E.; Koshland, D. E. *Science* **1993**, *262*, 1958.

(8) Yager, T. D.; Zewert, T. E.; Hood, L. E. *Acc. Chem. Res.* **1994**, *27*, 94−100. Smith, L. M. *Science* **1993**, *262*, 530−531. Cantor, C. R.; Broude, N.; Sano, T.; Przetakiewicz, M. *Mass Spectrometry in the Health and Life Sciences;* Burlingame, A. L., Carr, S. A., Eds.; Elsevier: New York, 1995; pp 513−532.

(9) Slater, G. W.; Mayer, P.; Drouin, G. *Electrophoresis* **1993**, *14*, 961−966.

(10) Laird, P. W.; Jackson-Grusby, L.; Fazeli, A.; Dickinson, S. L.; Jung, W. E.; Li, E.; Weinberg, R. A.; Jaenisch, R. *Cell* **1995**, *81*, 197−205.

(11) Marshall, E. *Science* **1995**, *268*, 1270−1271.

(12) Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *Anal. Chem.* **1990,** *62*, 882−899.

(13) Stults, J. T.; Marsters, J. C. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 359−363.

(14) (a) Nordhoff, E.; Ingendoh, A.; Cramer, R.; Overberg, A.; Stahl, B.; Karas, M.; Hillenkamp, F.; Crain, P. F. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 771−776. (b) Nordhoff, E.; Karas, M.; Cramer, R.; Hahner, S.; Hillenkamp, F.; Kirpekar, F.; Lezius, A.; Muth, J.; Meier, C.; Engels, J. W. *J. Mass Spectrom.* **1995**, *30*, 99−112.

(15) McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60−70. (b) McLuckey, S. A.; Habibi-Goudarzi, S. *J. Am. Chem. Soc.* **1993**, *115*, 12085−12095. (c) Habibi-Goudarzi, S.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 102−113. (d) Habibi-Goudarzi, S.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 102−113. (e) McLuckey, S. A.; Vaidyanathan, G.; Habibi-Goudarzi, S. *J. Mass Spectrom.* **1995**, *30*, 1222−1229.

(16) Potier, N.; Van Dorsselaer, A.; Cordier, Y.; Roch, O.; Bischoff, R. *Nucleic Acids Res.* **1994**, *22*, 3895−3903.

(17) Tang, K.; Taranenko, N. I.; Allman, S. L.; Chang, L. Y.; Chen, C. H. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 727−730.

(18) Little, D. P.; Chorush, R. A.; Speir, J. P.; Senko, M. W.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1994**, *116*, 4893−4897.

(19) Chen, R; Cheng, X.; Mitchell, D. W.; Hofstadler, S. A.; Wu, Q.; Rockwood, A. L.; Sherman, M. G.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 1159−1163.

(20) Zhu, L.; Parr, G. R.; Fitzgerald, M. C.; Nelson, C. M.; Smith, L. M. *J. Am. Chem. Soc.* **1995**, *117*, 6048−6056.

(21) Fitzgerald, M. C.; Smith, L. M. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 117−140.

(22) Little, D. P.; Thannhauser, T. W.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2318−2322.

(23) Greig, M.; Griffey, R. H. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 97−102.

(24) Little, D. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **1995**, *117*, 6783−6784.

S0002-7863(95)03361-0 CCC: $12.00

(M_r) by MS provides an independent verification, recognition of minor modifications (nine AT[4] and more than 10 p53[7] gene mutations are known) depends critically on mass accuracy and resolving power (RP). Further, molecular ion fragmentation (MS/MS) can produce sequence-specific product ions from oligonucleotides,[15,18] of special promise to identify and locate base changes of genetic defects. This is extended here to DNAs as large as a 108-mer.

Matrix-assisted laser desorption ionization (MALDI)[29] has yielded mass spectra[14,20,21] for DNA as large as a 500-mer.[17] However, sample dissociation has been a problem for many 30-mers and larger,[21] and enhanced instrumentation is necessary to resolve even the Na-adduct (+22 Da) peak of a 27-mer.[26] Similarly, special instrumentation made possible RP = 1150 for MALDI of a 12.4 kDa protein.[27] Sample dissociation is also a problem[22] for electrospray ionization (ESI)[30] of DNA, although M_r data for a 10^8 Da phage have been published.[19] Despite reports of M_r accuracies of 0.01%,[16] the same methods also can yield far poorer accuracy (~1%) due to unresolved impurities or adducts.[12] With conventional low-resolution instruments (RP 500−1000), the ubiquitous Na adduct is not resolvable for strands larger than ~35−70-mers.

This study exploits the combination[31] of ESI with Fourier transform (FT)[32] MS whose RP = 10^5 yields M_r values with <0.5 error for DNAs as large as 100-mers.[22] An earlier study showed that molecular ion dissociation with ESI-FTMS yields full sequence data for oligonucleotides ($n \leq 14$),[18] with the unique FTMS isotopic resolution especially important for ESI charge state assignment.[31b] However, only partial, and quite similar, fragmentation data were obtained for oligonucleotides as large as 25-mers by nozzle-skimmer (NS),[33] collisional activation (CA),[34] and infrared multiphoton dissociation (IRMPD).[35] Extending a recent Communication,[24] we report here that larger DNAs subjected to several fragmention methods can even yield complete sequence information, with spectra requiring only femtomoles of sample.

Experimental Section

Synthetic DNAs (sequences given in the relevant figures) from the Cornell Peptide/DNA Synthesis Facility or Perkin-Elmer Applied Biosystems (50-mer of Figure 3 and the 72- and 108-mers of Figure 5) were HPLC desalted as previously described.[22] For conventional ESI, ~100 mg of lyophilized DNA was dissolved in H_2O, sonicated, and diluted with CH_3CN to 4−20 mM (CH_3CN:H_2O, 3:1), with piperidine,[23] imidazole,[23] or triethylamine (TEA)[16] (0−0.02% by volume) added. Only negative ions were used; solutions were electrosprayed (−2.2 kV, coaxial SF_6 as electron scavenger gas)[36] at 1 mL/min. For ESI of nanoliter volumes,[28b,c] fused silica capillaries (10 mm id., Polymicro Technologies) were pulled and HF etched to a 2−5 mm tip i.d., suspended in triethylammonium acetate for 1 h, dried with a flow of high purity N_2, and Au coated. Solutions were prepared as for conventional ESI, but with higher H_2O (≤65%) and TEA or piperidine (≤0.1%) content. After tip loading by capillary action, the ESI voltage was ramped to ~700 V until the ion current stabilized; no pressure was applied for solution flow.[28a] The modified Finnigan FTMS instrument has been described in detail elsewhere.[31c] Briefly, the negative ions are transported through a heated capillary interface[37] and three rf-only quadrupole rods to an open cylindrical trap in a 6 T magnetic field, with trapping aided by a pulse of $\sim 10^{-6}$ Torr of N_2. A 30−60 s delay precedes broad-band excitation and detection, with 256K or 512K data points collected.

For NS dissociation,[33] the skimmer was held at −13 V and the capillary varied from −70 to −200 V, with a tube lens connected to the capillary adjusted from −200 to −300 V to maximize ion transmission. Despite an extensive investigation, experimental parameters affecting the production of b, c, d, and r ions are not well understood; repeat experiments after several months gave much lower abundances. For IRMPD,[35] the beam from a cw 27 W Synrad (duty cycle 10−50%) laser passed through a BaF_2 window along the magnetic axis. Except for same day experiments, dissociation efficiency is highly dependent on the exact focus position, so that 10−450 ms irradiation times were necessary to achieve the results shown. Extensive noncovalent adduction of the 100-mer was minimized with IRMPD (supporting information, Figure 3).[38,39] For multiple collisional activation (MECA),[40] trapped ions were subjected to ~100 highly attenuated broad-band chirp excitations (~4 V_{pp} [400 V_{pp} at 40 dB], 350 Hz/ms) separated by 5−10 ms, with N_2 pulsed again to $\sim 10^{-6}$ Torr. For single-frequency sustained off-resonance irradiation (SORI),[41] excitation at +1 and −1 kHz (separate spectra) relative to the precursor frequency was effected for 0.2 s.

Computer programs to determine fragment type and respective base composition (Table 1), to identify series of fragments differing by nucleotide units, and to find possible locations of these in a postulated sequence were written in BASIC and run on a 486 PC;[42] charge state assignment software[43] and deconvolutions to combine charge states of the same mass[44,45] were run on a Sparc 10 Sun workstation using PV-Wave. Mass (m) values are corrected for the number of negative charges by adding the mass of an equivalent number of protons. The reported m value is that of the most abundant $^{13}C_n$ isotopic peak, with n (except for $^{13}C_0$) denoted as an italicized integer separated from the m value by a hyphen. The most abundant isotopic peak is determined[46] from the deconvoluted abundances of all charge states; however, its fractional mass is an average of the values from all charge states.

(25) Barry, J. P.; Vouros, P.; Van Schepdael, A.; Law, S.-J. *J. Mass Spectrom.* **1995**, *30*, 993−1006.
(26) Christian, N. P.; Colby, S. M.; Giver, L.; Houston, C. T.; Arnold, R. J.; Ellington, A. D.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1061−1066.
(27) Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044−1050.
(28) (a) Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167−180. (b) Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. *Anal. Chem.* **1995**, *67*, 3802−3805. (c) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W. *Science* **1996**, *273*, 1199−1202.
(29) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53−68.
(30) Clegg, G. A.; Dole, M. *Biopolymers* **1971**, *10*, 821−826. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.
(31) (a) Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *85*, 9075−9078. (b) Henry, K. D.; McLafferty, F. W. *Org. Mass Spectrom.* **1990**, *25*, 490−492. (c) Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M. III; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 557−565. (d) McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379−386.
(32) Marshall, A. G.; Comisarow, M. B. *Chem. Phys. Lett.* **1974**, *25*, 282−283. Marshall, A. G.; Grosshans, P. B. *Anal. Chem.* **1991**, *63*, 215A−229A.
(33) Loo, J. A.; Udseth, H. R.; Smith, R. D. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 207−210.
(34) Carlin, T. J.; Freiser, B. S. *Anal. Chem.* **1983**, *55*, 571−574.
(35) Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B., McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809−2815.
(36) Wampler, F. M.; Blades, A. T.; Kebarle, P. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 289−295.
(37) Chowdury, S. K.; Katta, V.; Chait, B. T. *Rapid Commun. Mass Spectrom.* **1990**, *4*, 81−87.
(38) Little, D. P.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 209−210.
(39) Speir, J. P.; Senko, M. W.; Little, D. P.; Loo, J. A.; McLafferty, F. W. *J. Mass Spectrom.* **1995**, *30*, 39−42.
(40) Burnier, R. C.; Cody, R. B.; Freiser, B. S. *J. Am. Chem. Soc.* **1982**, *104*, 7436−7441.
(41) Gauthier, J. W.; Trautman, T. R.; Jacobson, D. B. *Anal. Chim. Acta.* **1991**, *246*, 211−225. Senko, M. W.; Speir, J. P.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2801−2809.
(42) Little, D. P. Ph.D. Thesis, Cornell University, 1995. Software can be obtained by sending a 1.44 M floppy disk to the corresponding author.
(43) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 52−56.
(44) Mann, M.; Meng, C. K.; Fenn, J. B. *Anal. Chem.* **1989**, *61*, 1702−1708. Reinhold, B. B.; Reinhold, V. N. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 207−215.
(45) O'Connor, P. B.; McLafferty, F. W. *J. Am. Chem. Soc.* **1995**, *117*, 12826−12831.
(46) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229−233.

Table 1. DNA Fragment Ion Types and Masses (Da)

sequence units	base (B) A	T	C	G
base (B), substituent[a]	134.047	125.035	110.035	150.042
nucleotide units: (B + C_5H_7O + PO_4H)	313.057	304.046	289.046	329.052
3′-end fragments				
w_n:[b] $(unit)_n$ + H + OH	331.068	322.056	307.057	347.063
5′-end fragments				
$\underline{a}_n$:[b,c] $(unit)_n$ + HO + C_5H_5O	411.094	402.082	387.083	427.089
$\overline{b}_n$:[d] a_n − C_5H_5O − PO_3H + H	251.102	242.090	227.090	267.097
c_n:[d] a_n − C_5H_5O − O + H	315.073	306.061	291.062	331.068
d_n:[d] a_n − C_5H_5O + H	331.068	322.056	307.057	347.063
r_n:[d] a_n − C_3H + H	375.094	366.082	351.083	391.089
internal fragments				
$\underline{i}_n$:[b] $(unit)_n$ + H + PO_4H + C_5H_5O	491.060	482.049	467.049	507.055
$\underline{i}_n$ − PO_3H[e] $(unit)_n$ + HO + C_5H_5O	411.094	402.082	387.083	427.089

[a] The base molecule is BH. [b] Ions can also be formed by loss of BH from this fragment ion. [c] McLuckey nomenclature is a_{n+1} − base.[15] [d] Observed only in NS spectra. [e] Observed only in MECA spectra.

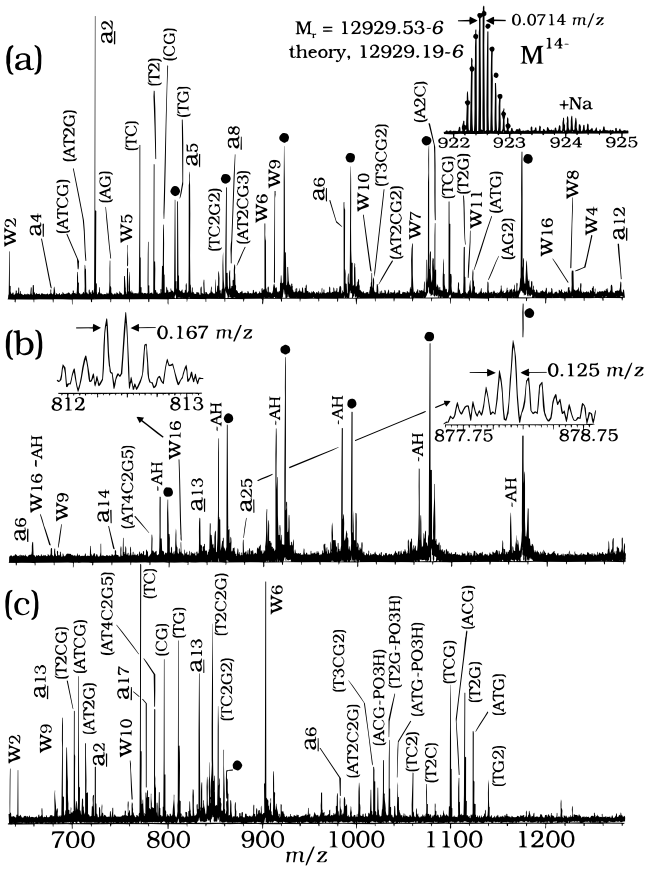


Figure 1. ESI/FTMS spectra of 42-mer of Figure 2: (a) NS, (b) IRMPD, (c) MECA. Top inset: expanded $(M - 14H^+)^{14-}$ region; small dots are best fit of theoretical isotopic abundances. ●: molecular ions. Fragment ion designations in text (in parentheses, compositions of internal ions).

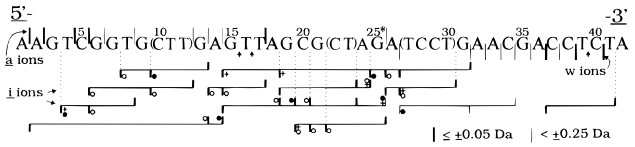


Figure 2. Sequence of 42-mer treated as an unknown: vertical bars up, cleavage yielding fragment ion with bases toward 5′-end (e.g., a ions); vertical bars down, fragment with bases toward 3′-end (e.g., w ions). Horizontal lines: internal (i) ions. Dark bars: mass error ≤±0.05 Da, average ±0.02. Light bars: mass error <±0.25 Da, average ±0.11. Asterisk and cross: unique to IRMPD and MECA, respectively. Small up arrow: sequence implied by lack of T loss; dotted vertical line, implied by i ions. Filled circle: other assignment(s) would involve T loss. Open circle: mechanistically most logical of assignments possible for this mass. Other internal ions observed have multiple sequence assignments. Parentheses in sequence: unknown order.

Results and Discussion

As examples, ESI/FTMS spectra of a 42-mer and a 50-mer DNA are shown in Figures 1 and 3. These provide mass values of 95 and 75 different fragment ions, respectively (Table 2 and supporting information, Table 1). However, the mass accuracy achieved (shown in Figures 2 and 4; dark bars show ≤0.05 Da error) makes quite specific sequence assignments possible for most peaks. The mechanisms providing these assignments will be discussed first.

Dissociation Mechanisms of Multiply Charged Anions. Previous studies[14b,15,18] of oligonucleotides have defined several fragmentation pathways; these and additional mechanisms found here for larger DNAs are summarized in Table 1 and illustrated in Scheme 1. A base unit (U) contains a base (B), a sugar, and a phosphate, designated by numbering from the 5′-end. Using the fragmentation nomenclature of McLuckey,[15] the 3′-end w and the 5′-end a − $B_{a+1}H$ fragment ions result from (Scheme 1) cleavage of a C−O bond between the 3′-deoxyribose and the phosphate with H-atom transfer from the sugar to the oxygen, triggered by the loss of the adjacent base B_{a+1} with a rearranged hydrogen atom[15,18,47] (loss of the B^- anion is also possible,[15e] but this yields the same corrected mass value as BH loss).

The mass of the a − $B_{a+1}H$ fragment ion represents sequence information for the 5′-bases only through base a; here this will be called instead the $\underline{a}_a$ fragment ion for clarity in spectral interpretation. The spectra of nine DNAs (42−108-mers, Figures 1−5) show 132 w and a peaks; 48% are formed by loss of base AH, 25% CH, 24% GH, and only 3% TH. These TH losses are only from the largest DNAs, nL-ESI 108-mer data and ^{5}TH in the 100-mer (Figure 5f,g). This T is adjacent to a ^{4}T base, which can also account for all of the w and a formation observed in the poly-T_{30} and poly-T_{60}. This negligible TH-loss tendency is useful in sequencing. If the exact mass

(47) Peaks designated here as w were previously termed[18] w + H in recognition of this rearrangement. The loss of the B^- anion is not differentiated here from the loss of BH; both give the same corrected *m* values. Contrary to previous reports,[15e,48] the spectra here show little evidence for B^- loss; selecting one charge state of a 14-mer and of a 35-mer for SORI dissociation showed only the loss of neutral base, not B^-. SORI dissociation of $(M - 12H^+)^{12-}$ of poly-T_{30} did give substantial T^- loss, but less than that of neutral TH; T is the least basic of the four bases.[49]

(48) Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. *Int. J. Mass Spectrom. Ion Processes* **1994**, *137*, 121−149.

(49) Liguori, A.; Napoli, A.; Sindona, G. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 89−93.

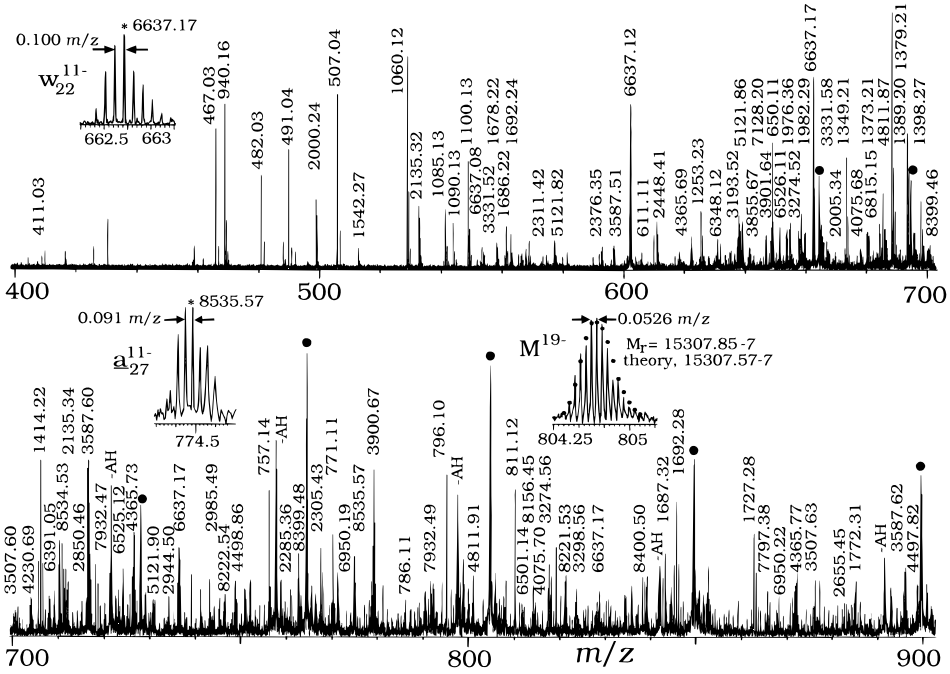


Figure 3. Partial IRMPD (10 ms) spectrum of 50-mer. Insets: isotopic distributions from w_{22}^{10-}, a_{27}^{11-}, and $(M - 19H^+)^{19-}$.

Table 2. Fragment Ion Masses of 50-mer DNA used in Figure 4, with Values from Different Charge States and/or Experiments Averaged

a ions: 411.06,[a,b,c] $\underline{a}_1$; 740.13,[b,c] $\underline{a}_2$; 1069.17,[b,c] $\underline{a}_3$; 1398.21,[b,c] $\underline{a}_4$; 1687.29,[a,b,c] $\underline{a}_5$; 1976.33,[a,b,c] $\underline{a}_6$; 2305.37,[a,b,c] $\underline{a}_7$; 2619.43-*1*,[b] $\underline{a}_8$; 2948.41-*1*,[b] $\underline{a}_9$; 3566.56-*1*,[b] $\underline{a}_{11}$; 3855.71-*1*,[a] $\underline{a}_{12}$; 4498.78-*2*,[a,b] $\underline{a}_{14}$; 4811.88-*2*,[a] $\underline{a}_{15}$; 7932.49-*3*,[a] $\underline{a}_{25}$; 8222.53-*4*,[a] $\underline{a}_{26}$; 8535.55-*4*,[a] $\underline{a}_{27}$

w ions: 611.10,[a,b,d] w_2; 940.15,[a,b,d] w_3; 1253.21,[a,b,d] w_4; 1542.23,[a,d] w_5; 2135.35,[a,b,d] w_7; 2448.40,[a,b] w_8; 2738.43-*1*,[b] w_9; 3331.52-*1*,[a,b,d] w_{11}; 4238.62-*1*,[b] w_{14}; 5121.86-*2*,[a,b,d] w_{17}; 6348.12-*3*,[a,b] w_{21}; 6637.13-*3*,[a,b] w_{22}; 6950.18-*3*,[a] w_{23}

internal ions: 1075.13,[d] $\underline{i}_{29-31}$; 1090.13,[a,b] $\underline{i}_{30-32}$; 1364.16,[b] $\underline{i}_{21-24}$; 1379.21,[a,d] $\underline{i}_{29-32}$; 1444.18,[b] $\underline{i}_{18-21}$; 1453.22,[b] $\underline{i}_{16-19}$; 1638.25,[b,d] $\underline{i}_{38-42}$; 1692.25,[a,b] $\underline{i}_{28-32}$; 1707.23,[b] $\underline{i}_{30-34}$; 1726.24,[b] $\underline{i}_{24-28}$; 1733.23,[b] $\underline{i}_{18-22}$; 1773.24,[b] $\underline{i}_{17-21}$; 1982.29,[a,b] $\underline{i}_{21-26}$; 2005.30,[a,b] $\underline{i}_{27-32}$; 2062.27,[b] $\underline{i}_{17-22}$; 2079.27,[b] $\underline{i}_{13-18}$; 2271.34,[b] $\underline{i}_{34-40}$; 2311.32,[a,b] $\underline{i}_{20-26}$; 2375.37,[a,b] $\underline{i}_{16-22}$; 2574.32,[b] $\underline{i}_{29-36}$; 2598.39,[b] $\underline{i}_{28-35}$; 2834.38-*1*,[b] $\underline{i}_{38-46}$; 2850.40-*1*,[b,d] $\underline{i}_{34-42}$; 2945.48-*1*,[a,b] $\underline{i}_{18-26}$; 2985.42-*1*,[b] $\underline{i}_{17-25}$; 3193.49-*1*,[a,b,d] $\underline{i}_{29-38}$; 3274.52-*1*,[a,b] $\underline{i}_{17-26}$; 3298.54-*1*,[a,b] $\underline{i}_{16-25}$; 3506.63-*1*,[a] $\underline{i}_{28-38}$; 4229.66-*1*,[a,b] $\underline{i}_{14-26}$; 4253.67-*1*,[b] $\underline{i}_{13-25}$; 4365.69-*2*,[a,d] $\underline{i}_{29-42}$

b, c, d, and r ions from NS: 1033.12, r_3; 1238.22, b_4; 1527.25, b_5; 1816.28, b_6; 1880.22, c_6; 1896.26, d_6; 2145.33, b_7; 2209.26, c_7; 2225.29, d_7; 2269.33, r_7; 2459.37-*1*, b_8; 2523.36-*1*, c_8; 2539.34-*1*, d_8; 2583.37-*1*, r_8; 2788.44-*1*, b_9; 2852.40-*1*, c_9; 2868.40-*1*, d_9; 2912.40-*1*, r_9; 3077.46-*1*, b_{10}; 3141.43-*1*, c_{10}; 33157.41-*1*, d_{10}; 3201.45-*1*, r_{10}; 3406.50-*1*, b_{11}; 3486.45-*1*, d_{11}; 3530.49-*1*, r_{11}; 3695.53-*1*, b_{12}; 3759.50-*1*, c_{12}; 3775.49-*1*, d_{12}; 3819.53-*1*, r_{12}; 4008.56-*1*, b_{13}; 4072.54-*1*, c_{13}; 4088.53-*1*, d_{13}; 4132.56-*1*, r_{13}; 4338.60-*2*, b_{14}; 4402.57-*2*, c_{14}; 4418.58-*2*, d_{14}; 4462.58-*2*, r_{14}; 4651.64-*2*, b_{15}; 4715.63-*2*, c_{15}; 4731.60-*2*, d_{15}; 4775.64-*2*, r_{15}; 4964.69-*2*, b_{16}; 5028.64-*2*, c_{16}; 5044.66-*2*, d_{16}; 5088.73-*2*, r_{16}; 5293.72-*2*, b_{17}; 5417.72-*2*, r_{17}; 5677.75-*2*, d_{18}; 5721.74-*2*, r_{18}; 6007.78-*3*, d_{19}; 6380.83-*3*, r_{20}

[a] IRMPD (10 ms) of molecular ions. [b] IRMPD (13 ms) of molecular ions. [c] NS of molecular ions. [d] IRMPD of w_{22} ions.

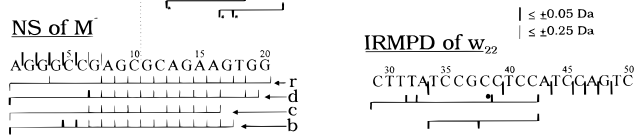


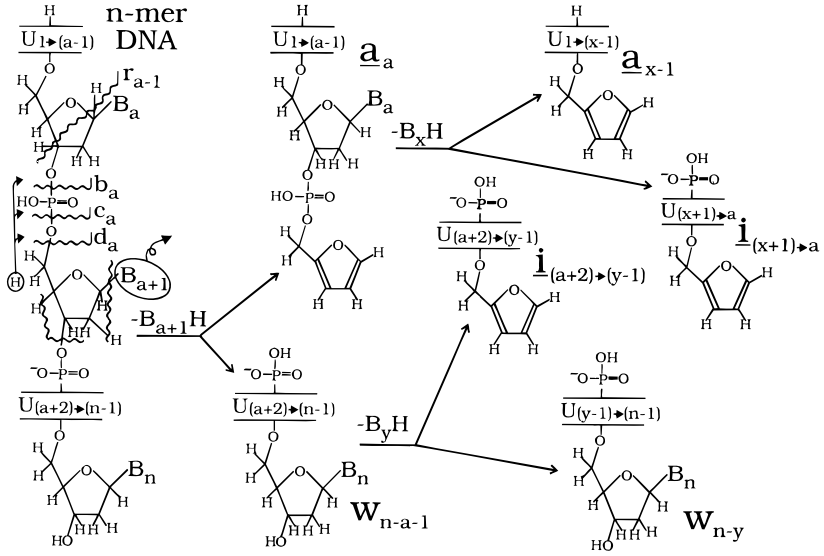
Figure 4. MS data providing sequence of 50-mer treated as an unknown (symbols as in Figure 2, and mass values in Table 2). Other internal ions for which multiple sequence assigments are possible are listed in the supporting information, Table 1. Asterisk: unique to 13 ms IRMPD. Dark (light) bars: mass error ≤±0.05 Da, average ±0.02 (≤±0.25 Da, average, ±0.12).

difference between two fragment ions corresponds to T and one (or more) other base(s), the base lost in forming the lower mass fragment ion is *not* T unless the fragment ion contains an adjacent T.

At lower energies, the loss of base AH, CH, or GH can occur without backbone cleavage,[15,18] providing no sequence information. For example, the IRMPD spectrum of the 42-mer (Figure 1b) shows minor losses from M^{15-} of AH, CH, 2AH, (AH + CH), and GH, in order of abundance, while the MECA spectrum shows similar base losses, but only for M^{15-}. The SORI spectrum of $\underline{a}_{13}^{5-}$ from the 42-mer (supporting information, Figure 1) shows $\underline{i}_{2-13}$ (see below) and base losses AH > GH > CH, with no TH loss. Further loss of a neutral base (AH > CH, GH ≫ TH) has been observed to occur from w, a, and internal fragments (Tables 2 and 3). Ions corresponding to the loss of a PO_3H from internal ions are also occasionally observed (e.g., MECA spectrum, Figure 1c).

The larger DNAs examined here show a previously unrecognized mechanism valuable for sequencing. Internal (i) ions can be produced by secondary fragmentation (with neutral base loss) of either the w or the a fragments, accompanied by formation of smaller w or a fragment ions, e.g., $\underline{a}_a \rightarrow B_xH + \underline{a}_{x-1} + \underline{i}_{(x+1)\rightarrow a}$ (Scheme 1). Of the 110 flanking bases lost informing unambiguously assigned internal ions in these DNA spectra, 57% are A; i ions are formed by T loss at the $^{30}TT^{32}T$ group in the 50-mer, in the IRMPD spectra of its w_{22}, of poly-T_{30}, and of the T-rich 25-mer $AT_2AT_3AT_4AT_5AT_6$.[18] To illustrate the utility of i ions, for the 42-mer (Figure 2), identifying $\underline{a}_2$ does not tell what base (^{3}GH) is lost in its formation, although $\underline{a}_4$ shows that bases 3 and 4 are (TG) (parentheses indicate unknown base order); the complement w_{39} formed from the same dissociation of the molecular ion is presumably too unstable to be observed. However, an internal ion of the composition T_2CG_2 can only represent bases 4−8

Scheme 1



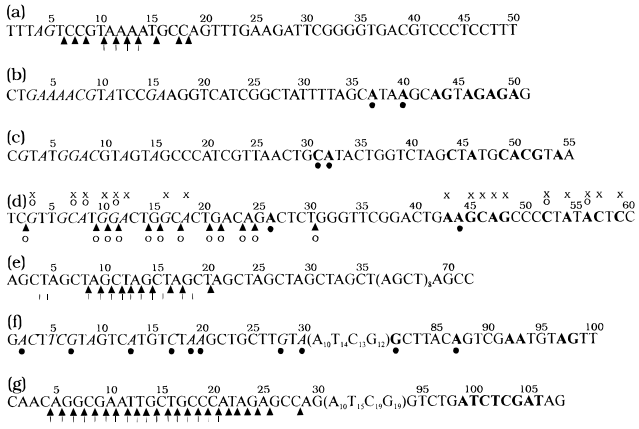


Figure 5. ESI/FTMS spectra of DNAs: (a) another 50-mer, NS; (b) 51-mer, NS and IRMPD; (c) 55-mer, NS and IRMPD; (d) 60-mer, NS (O: −105 V, X: −130 V) and IRMPD; (e) 72-mer, NS; (f) 100-mer, NS and IRMPD (supporting information, Figures 3 and 4 and Table 2 show more extensive data and treatment as an unknown); and (g) 108-mer, NS and NS from nanoliter flow ESI (spectra not measured below *m/z* 600). Vertical lines and triangles below: observed b and d ions, respectively. Bold and italic: base lost to form w and a, respectively; dot below base, observed only by IRMPD.

($i_{9\rightarrow13}$ would require the loss of ^{8}T), ordering the bases ^{3}G^{4}T. Internal ions can also be formed by subjecting a product ion to further dissociation (MS^3), such as the IRMPD of w_{22}^{10-} from the 50-mer (Figure 4, lower right). This produces i fragments not found in IRMPD of M^- (Figure 4, top), including $i_{29\rightarrow31}$ formed by the loss of ^{32}T adjacent to ^{31}T.

For the ESI spectra of oligonucleotides,[15,18] NS dissociation[33] and IRMPD[35] yielded very similar product ions. However, the NS spectrum of the 50-mer (Figure 4) is dominated by entirely different fragment ions containing the 5′-end, designated here as b, c, d, and r. These are lower in mass vs the a peak by 159.99, 96.02, 80.03, and 36.00 Da, respectively, corresponding to losses of $C_5H_5O_4P$, $C_5H_4O_2$, C_5H_4O (HPO_3 is 79.97 Da), and C_3. The requirements for formation of these products are unclear (Figure 5). While the 51-, 55-, and 100-mers gave no b, c, d, or r ions, these ions provided sequence information for the two 50-mers and the 60-, 72-, and 108-mers from bond cleavages of 18, 10, 12, 14, and 23 5′-end bonds, respectively. NS spectra of the first 50-mer measured several months later under apparently similar conditions showed few such ions, while those of the 60-mer at higher energies using the same solution conditions generated no d ions, but instead gave both a and w sequence ions. The 108-mer (Figure 5g) also gave no b or d ions with a nanoliter volume ESI, and the d_n ions from normal (milliliter) volume ESI were 1.01 Da (-H) lighter than expected, e.g., the mass difference of d_{11} and b_{11} is 78.96 Da (3437.58-*1* − 3358.62-*1*) vs 79.96 Da expected. IR-MALDI of a singly-charged 19-mer $A_5T_4C_7G_3$ oligonucleotide also produces $d_{2-4,8}$ ions (termed "X" ions), but not b, c, or r fragments.[14b]

These new products appear to result (Scheme 1) from backbone cleavages at the O−P bond (b ions), P−O bond (c ions), O−ribose bond (d ions), and across two ribose bonds (r ions), with transfer *to* the 5′-product ion of one, one, one, and two H atoms, respectively (Table 1). In contrast, the backbone cleavage yielding the w and a ions transfers H *away* from the 5′-end. Possibly these new 5′-products arise from initial loss of the B^- anion (not BH) on the 3′-side of the products. If the base at this position is T or C (pyrimidine bases), the tendency for b, c, d, and r ion formation is smaller, but not as effective as the base T in reducing w, a, and i formation; note that the influential base is on the ribose from which the hydrogen(s) are transferred, irrespective of the backbone bond cleaved in forming these major products. This anion loss would decrease the negative charge on the 3′-end, providing a higher tendency for intramolecular transfer of H^+ toward the 5′-end to form b, c, d, and r ions by cleavage on the 5′-side of B^- loss. It is conceivable that the initial B^- anion displacement is effected by an *intermolecular* reaction;[24] no complementary 3′-end ions could be found for b, c, d, or r ion formation, as can be found for the other unimolecular dissociations of multiply charged ions. However, initial experiments on bimolecular NS reactions of these multiply charged DNA anions with several other anions, e.g., $(RO)_2PO_2^-$, gave no such products.

Methods for DNA Ion Dissociation. The earlier oligonucleotide sequencing[15,18] utilized NS,[33] collisional activation (CA),[34] and IRMPD[35] techniques. Here the FTMS CA methods of multiple excitation (MECA)[39] and sustained off-resonance irradiation (SORI)[40] were also employed, as was 193 nm laser photodissociation.[50] However, all methods were not tested under a comprehensive variety of experimental conditions, such as temperature, pressure, amount of energy added, and dis-

(50) Williams, E. R.; Furlong, J. J.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1990**, *1*, 288−294. Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little, D. P.; McLafferty, F. W. *Int. J. Mass Spectrom. Ion Processes*, in press.

sociation times. CA of an 8-mer under ostensibly similar conditions (although different types of mass spectrometers) gave similar spectra (w and a peaks) in two laboratories,[15,18] but a far more complex spectrum (w, a, x, y, z, and fragment − H_2O ions) in a third laboratory.[51] Similarly, in the present studies, the dissociation of larger nucleotide ions was also found to be sensitive to experimental conditions as well as to DNA structure (vide supra).

Using the 42-mer as an example (Figure 1), NS, IRMPD, and MECA data provide complementary information. NS yields no b, c, d, or r ions, as found for smaller DNAs,[18] but provides the most complete w and a series, as summarized in Figure 2. IRMPD provides the valuable complementary pair 7912.53-*3* + 4880.95-*2* + AH (135.05) = 12928.53-*5* (M_r = 12929.53-*6*). MECA yields the most internal ions. Initial trials with SORI (+1 kHz from resonance) of the 42-mer molecular ions (data not shown) gave primarily base loss ions, as did SORI (−1 kHz from resonance) of its $\underline{a}_{13}^{5-}$ fragment ion (supporting information, Figure 1); higher collisional energies gave poor product trapping efficiency under the conditions investigated. Figure 5b,c,d compares the sequence data from NS and IRMPD of the 51-, 55-, and 60-mers; each shows two w ions formed only by IRMPD. Limited spectra from 193 nm photodissociation[50] of the molecular ions of the Figure 4 50-mer and poly-T_{30} (data not included) show extensive precursor ion depletion but only low intensity product ions; poly-T_{30} yielded no ions that were necessarily formed by T loss, such as $w_{2-17,29}$ and product ions complementary to $w_{8,11,14,15,23}$ without T loss.

These studies, although not comprehensive, led to the selection of IRMPD as the primary method for sequencing. Absorption of an IR photon adds a specific energy value, so that IRMPD should deposit a narrow distribution of energy values; covalent bond dissociation is minimal during IRMPD "boiling off" of non-covalent adducts.[38,39] NS dissociation has the unique advantage of producing the 5′-fragments b, c, d, and r for some larger nucleotides, but the precursor ion cannot be selected for MS/MS.

Sequencing Strategy. For large proteins (e.g., 29 kDa), a "top down" approach yields MS^n sequence information without prior degradation of the protein.[31d,39,52,53] This strategy was used in part for the oligonucleotide ESI/FTMS data;[18] its adaptation here for the DNA spectra involves the following steps:

(1) Assign exact mass (*m*) values for the most abundant isotopic peak of the molecular ion (M^-) and of each fragment ion, averaging all charge states.

(2) Identify complementary ions, those whose *m* values sum to that of a precursor ion minus that of the base lost (favored: AH > CH, GH ≫ TH) with each bond cleaved. Identify (a) masses of fragment pairs plus a BH mass whose sum equals the mass of the molecular ion; (b) repeat this for higher multiplets (terminal and internal ions) whose *m* values were not already used in a previous ("unique") complementary set; and (c) find complementary sets for the larger fragment ions.

(3) For each *m* value: (a) use exact masses to identify the fragment ion type(s) (e.g., w, a, i, Table 1) and combination(s) of the four bases that match the *m* value within experimental error; (b) identify neighboring fragment ions whose difference in base assignments (or *m* values) can correspond to a specific base unit or simple combination of base units; and (c) identify contiguous assignments of such base units which should thus represent a partial sequence series in the molecule.

(4) Construct a trial sequence(s) by placing these assignments within the overall molecular weight restriction: (a) place the 3′- and 5′-data (w and a ions) at the corresponding ends of the molecule; (b) place any complementary doublet from step 2 that contains a 3′- or 5′-assignment and assign its 5′- or 3′-counterpart; (c) place the remaining ion series similarly; (d) place all other fragments (larger a and w plus i) within this trial sequence; (e) cross-check for the most probable combination of sequence assignments, including the base loss probability A > C, G ≫ T [If the exact mass difference between two fragment ions corresponds to T and one (or more) other base-(s), the base lost in forming the lower mass fragment ion is *not* T (unless adjacent to a T)]; and (f) list all peak *m* values that are unassignable in the trial sequence.

(5) Perform additional experiments: (a) molecular ion fragmentation using different energies or methods and (b) MS^n of specific fragment ions for localized sequence information.

Mass Accuracy Restrictions on Base Assignments. In applying this strategy, the degree of confidence in the assignments of strategy steps 2 and 3, especially of fragment types and base composition, is critically dependent on mass accuracy. Only a restricted number of mass values are possible for combinations of the four bases and the Table 1 fragment ion types.[18] For example, the mass of an a vs that of an i ion will be heavier by only 0.046 Da for a composition difference of $-C_2$, $+G_2$, such as $\underline{a}_5$-ACG_3, 1687.297 Da vs $\underline{i}_5$-AC_3G, 1687.251 Da. Thus when three spectra of a 50-mer DNA gave peaks with an average mass of 1687.29 Da (Table 2), they were initially assigned as $\underline{a}_5$. The measured masses of the larger a ions in Table 2 also agree better with those expected for $\underline{a}_{6-9}$, $\underline{a}_{11}$, $\underline{a}_{14}$, and $\underline{a}_{15}$ (<0.05 Da error, Figure 4) than for the isobaric i ions, especially after correcting for calibration errors of $+15 \pm 3$ ppm, -34 ± 7 ppm, and -9 ± 4 ppm (internal calibration) in each of the three spectra. However, only external frequency calibration was used for the data reported here. Other checks can also make such distinctions. In the strategy step 3a above, the NS fragments b_{15}, c_{15}, d_{15}, and r_{15} ions support the IRMPD 4811.88-*2* assignment as an $\underline{a}_{15}$ ion, not the $\underline{i}_{15}$ ion $A_4C_6G_5$, 4811.774-*2* or $A_4T_8CG_2$, 4811.756-*2*; for the two possible a_{15} compositions, $A_4C_4G_7$, 4811.820-*2* is preferable to $A_9C_2G_4$, 4811.857-*2* based on the lower A assignment (Figure 4, top) of 2305.44 and 3855.71-*1* as the $\underline{a}_7$-AC_2G_4 and $\underline{a}_{12}$-$A_2C_4G_6$, respectively. Similarly, for the IRMPD 3331.55-*1* fragment, identification of most of the w_2 through w_8-$A_2T_2C_3G$ sequence peaks favors w_{11}-$A_2T_3C_5G$, 3331.549-*1* over $\underline{i}_{10}$-$A_3T_2CG_4$, 3331.543-*1*. Figures 2 and 4 and the supplementary information show that the majority of assignments are within ±0.05 Da of the correct value.

Sequencing Unknown DNAs. Although the 50-mer sequence was known, its ESI/FTMS data will be treated here as an unknown to test the proposed top-down sequencing strategy. For strategy step 1, the data from the "soft" ESI spectrum yielding M^- ions and from the IRMPD MS/MS spectra at two irradiation levels give M_r = 15307.85-*7* and the 60 IRMPD fragment ion masses listed in Table 2. For the complementary ions of step 2 (Figure 4, upper vertical a line separated from lower vertical w line by the base lost), the *m* values of two pairs plus that of base AH sum to that of M_r: 8222.53-*4* + 135.05 + 6950.18-*3* = 15307.76-*7* (actually $\underline{a}_{26} + {}^{27}A + w_{23}$) and 8535.55-*4* + 135.05 + 6637.17-*3* = 15307.77-*7* (actually $\underline{a}_{27} + {}^{28}A + w_{22}$). Both a and w primary fragments also exhibit complementary pairs of a or w ions, respectively, with an i ion, separated by the base lost in their formation. Referring to the

(51) Pomerantz, S. C.; McCloskey, J. A. Proceedings of the ASMS Conference on Mass Spectrometry and Allied Topics, Atlanta, GA, May 1995, p 600.

(52) O'Connor, P. B.; Speir, J. P.; Senko, M. W.; Little, D. P.; McLafferty, F. W. *J. Mass Spectrom.* **1995**, *30*, 88−93.

(53) Aaserud, D. J.; Little, D. P.; O'Connor, P. B.; McLafferty, F. W. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 871−876.

top set of horizontal i lines (Figure 4), for 8222.53-*4*, 3855.71-*1* + 135.05 + 4229.66-*1* = 8220.42-*2* (actually a_{12} + ^{13}A + i_{14-26}), and for 6950.18-*3*, 2598.39 + 111.04 + 4238.62-*1* = 6948.05-*1* (actually i_{12-35} + ^{36}C + w_{14}). These and other large fragment ions yield additional pairs for a total of 13 complementary sets in Figure 4. Other observed i ions of multiple possible assignments are listed in the supplementary information.

For strategy step 3, the masses of the fragment ions shown along the top of Figure 4 agreed best with assignments as a and w ions. The corresponding base compositions are matched in step 3b to find simple differences (and confirm the assignments), and these fragments have been arranged as partial sequence series for the w, a, and i data. Though there are several assignments for the 4811.88-*2* ion, its difference from the 4498.78-*2* ion of 313.10 Da indicates these to be the same type of fragment but differing by an A unit (313.06). Repeating the process finds smaller ions of the same type differing by (AG), C, (CG), G, A, G, C, C, G, G, G, and A, with the smaller ones certainly of the a type. This process identifies most a ions through a_{15} and many w ions through w_{23}.

For step 4, these a and w ion placements (Figure 4, top) are used to assign sequence positions to the complementary ions. The 8222.53-*4* and 8555.35-*4* must represent a fragments (actually a_{26} and a_{27}), as their complements have been identified as w_{23} and w_{22}. The next smaller a fragment (7932.49-*3*) is then identified by the mass difference 290.04-*1* (base C unit, 290.05-*1*). The i ion 4229.66-*1* found above to be complementary with a_{12} and ^{13}A now can be placed between these a fragments (actually i_{14-26}). The 1982.29 Da ion should be i-T_2C_3G, with exact mass differences identifying its neighbors i-$T_2C_3G_2$, i-$T_3C_3G_3$, and i-$T_3C_3G_4$. The latter's mass difference vs the 4229.66-*1* ion identifies it as $A_2T_3C_3G_5$, consistent with the a_{14}, a_{15} identification of ^{15}A, and establishing ^{14}G. Other i ions are now uniquely assignable to the base composition possibilities found only in this region. For example, now the 1444.18 (T_2G_2) can only be i_{18-21}, and 1364.16 (T_2C_2) can only be i_{21-24}. Similar assignments of complementary ions at the 3′-end are unique, extending the sequence information.

For step 4e, the majority of postulated fragmentations have resulted from loss of base A, and the only fragment ion from T loss is the 786.11 (T_2), explained by the presence of three adjacent T bases, ^{30}T−^{32}T. The expected high stability of the 3′-bond of a T-containing ribose is valuable in indicating the correct order for the pairs ^{18}TG, ^{21}TC, and ^{44}TC; in each case, a base other than T is lost in forming the smaller fragment ion defining the pair. The ordering of ^{49}TC is indicated by the absence of a w_1 fragment (as in Figure 2), but the specificity of this correlation should be checked with further examples. None of the 138 fragments has unexplainable masses (Table 2 and supplementary information), and no other sequence was found that fit the data; an algorithm to check this exhaustively is under development. Thus two IRMPD MS^2 spectra correctly show all the sequence except the region $^{10-11}$(CG). A third IRMPD spectrum (not shown) also produced a 2064.28 fragment corresponding (error, −0.05; no T loss) to A_3CG_2, i_{11-16} that defines these final positions ^{10}C^{11}G. For step 5a, the value of additional data from a higher energy IRMPD MS^2 experiment (spectrum not shown) is indicated by the asterisks in Figure 4. NS fragmentation yields the additional (Figure 4, lower left) b, c, d, and r fragments that provide the correct ordering of ^{10}C and ^{11}G to complete the sequence. The NS spectrum also extensively confirms the remainder of the 5′-sequence through ^{20}G. To confirm the 3′-sequence (step 5b), the abundant IRMPD w_{22} peak (6637.17-*3* Da) was further dissociated by IRMPD (MS^3, Figure 4, lower right). Its identifiable 5′-end fragment ions correspond to cleavages of 12 of its 21 bonds, completely supporting the sequence postulated from the IRMPD MS^2 spectra.

The 42-mer DNA, when subjected to NS (no b, c, d, or r ions), IRMPD, and MECA (Figure 1) gave data providing the sequence information of Figure 2, again treating this as an unknown (measured M_r = 12929.53-*6*). The complementary M^- pair 7912.53-*3* and 4880.95-*2* (+ AH = 12928.53-*5*; actually a_{25} and w_{16}) can be readily identified, as the latter is shown to be the w fragment by the complementarity w_{10} (3051.62-*1*) + 1693.25 + AH (sum, 4879.92-*1*). Mass differences give assignments of high confidence for the majority of a ions through a_{17} and w ions through w_{16}. The other complements can be applied to this framework. A series of i ions with a common 3′-end can be established by mass differences: i_{15-31}, i_{19-31}, i_{25-31}, i_{26-31}, and i_{27-31}, positioned by complementarities such as a_{13} + AH + i_{15-31} + AH + w_{10}. Other i ions provide additional sequence delineation and confirmation. The (T_xB) ordering (step 4e) provides the sequences ^{15}GT^{17}T and ^{39}T^{40}C. These primary dissociation spectra (no MS^3) of the 42-mer, without b, c, d, or r peaks, yield the correct molecular base composition and base identities for positions 1−9, 13−21, 24−26, and 31−42.

Sequence data for seven other 50−108-mer DNAs are summarized in Figure 5. For the 100-mer, treating the NS and IRMPD data as an unknown (supplementary information, Table 2 and Figures 3 and 4) yields nearly complete sequence information for bases 1−29 (all but 10−12, even without b, c, d, and r ions) and partial information for 72−100. All the remaining peaks can be correlated with the known structure; 12 appear to be produced by base losses in the 30−71 base region. MS^3 spectra from further dissociation of fragment peaks (e.g., i_{31-46}, i_{48-74}) should check this and give additional sequence data.

Point Mutation Screening. A change in molecular weight is not only definitive evidence of a molecular modification, but the corresponding mass shift in a spectral fragment(s) could identify the structural change(s) and restrict its location. A 50-mer was synthesized to be the same as that of Figures 3 and 4, except to contain a mutation(s) unknown to the authors. Its −9.04 shift in the M_r value was consistent (Table 1) with that of A → T (Figure 6a); this could also be, for example, A → G combined with G → T. IRMPD of the M^- ions generated a spectrum (in <1 min) quite similar to Figure 3, including unshifted fragment masses for a_{14}, a_{25}, and w_{22}, constraining the mutation to bases 26−28 (T replacing ^{27}A or ^{28}A). The a_{27} fragment shifted by −9.08 (8535.57-*4* → 8526.49-*4*), pinpointing the mutation as ^{27}A → ^{27}T. Other evidence supports this: the Figure 4 w_{23} ion (formed by loss of ^{27}AH in the normal 50-mer) is not observed, as this requires ^{27}TH loss in the mutated strand.[24]

Sample Requirements for Sequence Verification. The recently developed ESI capability in which sub-nanoliter amounts of sub-micromolar protein solutions provide FTMS spectra[28c] has been applied to the 50-mer DNA of Figures 3 and 4. The Figure 7 spectrum produced from 8×10^{-16} mol introduced ($<10^{-16}$ mol consumed in the 3 s ion accumulation) has good signal/noise, yielding an M_r value with 0.2 error from a spectrum measured in less than 1 min. This 10^3 improved sensitivity vs MALDI,[14b] and classical sequencing methods for the ∼50-mer should prove valuable for applications such as gene-level diagnoses.

Conclusions

Molecular ion dissociation at present limits our applications to unknown 100-mer DNAs, but others have been successful

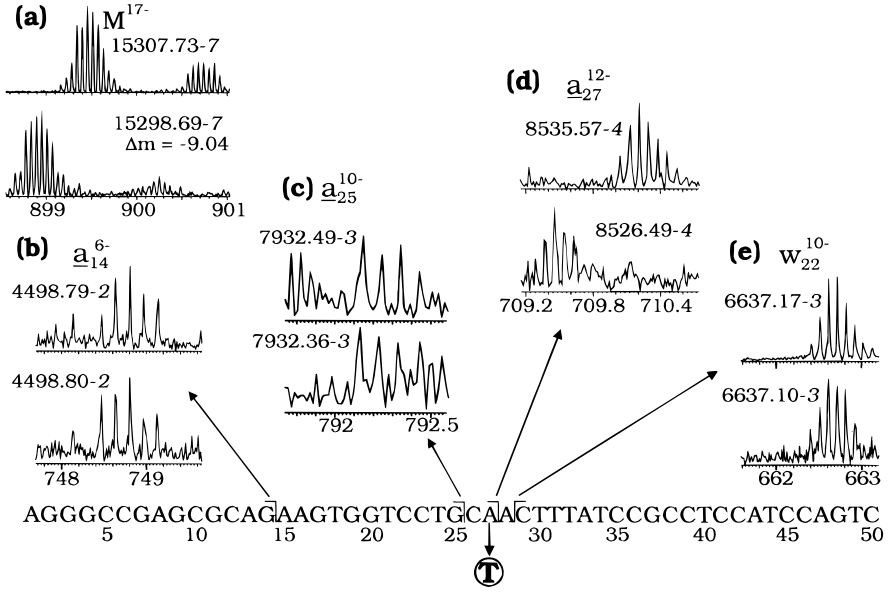


Figure 6. IRMPD spectra of 50-mer (top spectra) and its unknown $^{27}A \rightarrow {}^{27}T$ mutant (bottom): (a) molecular ions, (b) a_{14}^{6-}, (c) a_{25}^{10-}, (d) a_{27}^{12-}, and (e) w_{22}^{10-}.

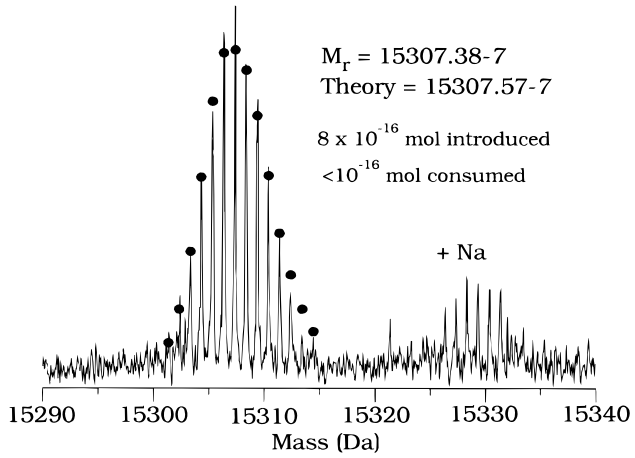


Figure 7. Deconvoluted (charge states summed) ESI/FTMS spectrum of 8×10^{-16} mol of the Figure 4 50-mer (1.5 nL of a 500 nM solution) Small dots: see Figure 1.

in volatilizing intact larger molecules.[16,17,19] Using Watson−Crick paired double-strand DNA may ameliorate this problem; for a 39 kDa double strand DNA, the molecular ion peaks are the most abundant in the spectrum.[54] The high-resolution, high mass accuracy spectra from ESI/FTMS generated by NS, IRMPD, and CA methods can even provide complete sequence information for an unknown 50-mer DNA. These measurements are complementary to conventional sequencing techniques; routine measurement times of ≪1 min are conceivable, and algorithms for efficient data reduction are under development. The sub-femtomole sample requirements for M_r data achieved here have since been reduced further by $\times 10^{-3}$ for proteins, with nine fragment ions of accurate mass also measured for 10^{-17} mol of a 29 kDa protein.[28c] Of special promise is the high-throughput screening for mutations in relatively complex mixtures of DNA molecules; component molecular ions of altered M_r values can be subjected to MS/MS to constrain the sequence location of the mutation.

Acknowledgment. The authors thank E. K. Fridriksson, Z. Guan, N. L. Kelleher, P. B. O'Connor, C. Van Pelt, and T. D. Wood for experimental assistance, T. P. Begley, C. R. Cantor, C. A. Costello, R. Nicewonger, T. Sano, and S. Taylor for helpful discussions, J. L. Peng, R. W. Sherwood, and T. W. Thannhauser for DNA synthesis and purification, A. Andrus for additional DNA synthesis, G. H. Morrison for picospray tip fabrication equipment, and the National Institutes of Health, Grant GM16609, for financial support.

Supporting Information Available: Tables giving fragment ion masses and identification from MS/MS spectra of 50-mer (fragments not in Figure 4 and Table 2) and 100-mer (Figure 5f) DNA and figures showing the SORI spectrum (MS^3) of the a_{13} of the 42-mer (from Figures 1 and 2), NS spectrum of 50-mer (Figures 3,4), IRMPD spectrum of the 100-mer (Figure 5f), and sequence of this 100-mer treated as an unknown (7 pages). See any current masthead page for ordering and Internet access instructions.

JA9533611

(54) Aaserud, D. J.; Kelleher, N. L.; Little, D. P.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* Accepted for publication.